Journal of Computer Science and Applications.
ISSN 2231-1270 Volume 6, Number 2 (2014), pp. 107-113

http://www.irphouse.com

Fortified Similarity Integration In Image-Rich Web Sites Using Simlearn

Muni Manasa G and S.M.Farooq

M.Tech, CSE Department,
Madanapalle Institute of Technology & Science,
JNTUA, Madanapalle, AP, Iindia.
munimanasa8@gmail.com
Asst. Professor, CSE Department,
Madanapalle Institute of Technology & Science,
JNTUA, Madanapalle, AP, Iindia.
farooqsm@mits.ac.in

Abstract

Digital images gained popularity due to increased usage of internet. People who are working on different domains are seeking for the opportunities offered by the capability to retrieve and manipulate remotely stored images in all possible ways. Not only social networking sites but also some online shopping sites are also furnishing product related pictures. Images in social networks are having marginal notes, remarks, and other information which forms a heterogeneous image-rich information networks. Retrieving information from huge image rich sites is very helpful but it is very hard and challenging job because those sites are having lot of data in the form of pictures, text, users, group and most importantly network structure. There is an algorithm namely k-SimRank (HMok-SimRank) to compute link-based similarity in weighted heterogeneous information networks, which is having problem with high cost and low efficiency. Another problem with the existing algorithm is that it won't get relevant image while retrieving. A new algorithm came into existence with the features of fast retrieval and data relevance namely SimLearn. More over text based and feature based similarities of heterogeneous images are also found.

Paper code: 14618-JCSA revised paper

I. Introduction

Retrieving images is a process of extracting images from large data base. To extract images from such data bases is very difficult and more time consuming process. By entering image related query user get the image based on the key words or query. Some of the image retrieval methods are

Content based image retrieval (CBIR):

The main concept of CBIR is instead of retrieving images based on textual query, retrieving the images based on features of images and pixels. This conventional technique has lot of drawbacks, like more time consuming and cost .For effective image retrieval we are using content-based image retrieval technique. Today's world user need to retrieve image from verity of domains based on their professions including crime avoidance, medicine, and fashion etc.

Link based similarity search:

Applications like social networks and biological databases having graphs like structure .we can understand the relationship between the nodes with the help of link prediction, recommendations and spam detection methods .with the help of link based similarity join we can find the link based similarities.

Not only social networking sites but also some online shopping sites are also furnishing product related pictures. Images in social networks are having marginal notes, remarks, and other information which forms a heterogeneous image-rich information networks. Retrieving information from huge image rich sites is very helpful but it is very hard and challenging job because those sites are having lot of data in the form of pictures, text, users, group and most importantly network structure. With the help of link based similarity methods we can retrieve images with information.

II. Related work

A. Scope of the project

Retrieving information from huge image rich sites is very helpful but it is very hard and challenging job because those sites are having lot of data in the form of pictures, text, users, group and most importantly network structure. There is an algorithm heterogeneous minimum order SimRank to compute link-based similarity in weighted heterogeneous information networks**.** To retrieve images from image rich information networks we are having algorithm namely mok- simRank. Existing algorithm have a problem with high cost, low efficiency and more time consuming process. A new algorithm came into existence namely Simlearn1. Simlearn having features of combining both link based and content based similarities.

A. Existing System:

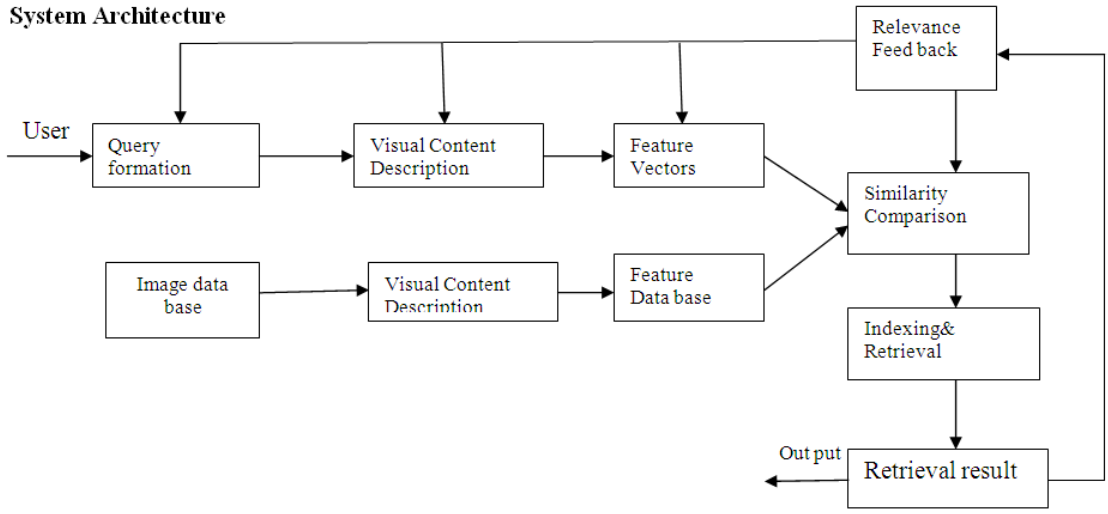
People who are working on different domains are seeking for the opportunities offered by the capability to retrieve and manipulate remotely stored images in all possible ways. Not only social networking sites but also some online shopping sites

are also furnishing product related pictures. Images in social networks are having marginal notes, remarks, and other information which forms a heterogeneous image-rich information networks. Retrieving information from huge image rich sites is very helpful but it is very hard and challenging job because those sites are having lot of data in the form of pictures, text, users, group and most importantly network structure. There is an algorithm heterogeneous minimum order SimRank to compute link-based similarity in weighted heterogeneous information networks**.**

III. Proposed work

Generally heterogeneous information net work having there type of nodes namely groups, tags, annotations.Simrank is a standout amongst the most famous calculations for evaluating item similarity in data systems. It computes the similarity between articles focused around the instinct that two items are comparative so they are connected by comparable queries in the system.

There are two impediments with Simrank: (1) It is expensive to figure and not adaptable to large datasets. (2) It measures object likeness singularly by connection data. However, in picture rich data systems, object likeness can likewise be evaluated by picture characteristic. To address the above two problems, we introduce an e-cient approach called MoK-SimRank to improve the speed of SimRank, and propose an algorithm called SimLearn to consider both link and content information by seamlessly integrating reinforcement learning with feature learning.



A.Query formation:

The images that are retrieved by user can be classified into different types. Commonly used query formations are: category browsing, query by concept, query by sketch, and query by example. Category browsing is to browse through the database according to the category of the image. For this purpose, images in the database are classified into diverse categories according to their semantic or visual content.

B .Feature vectors:
This extracts the features of images. The features may be color, texture, shape and spatial relationship.

C. Feature data base:
Featur**e** data base having the different features of images that generally images contain.

D. Similarity Measures and Indexing Schemes:
Instead of exact matching, content-based image retrieval calculates visual similarities between a query image and images in a database. Accordingly, the retrieval result is not a single image but a list of images ranked by their similarities with the query image. Many similarity measures have been developed for image retrieval based on empirical estimates of the distribution of features in recent years. Different similarity/distance measures will affect retrieval performances of an image retrieval system significantly.

E. **Indexing scheme**
Another important issue in content-based image retrieval is effective indexing and fast searching of images based on visual features. Because the feature vectors of images tend to have high dimensionality and therefore are not well suited to traditional indexing structures, dimension reduction is usually used before setting up an efficient indexing scheme.

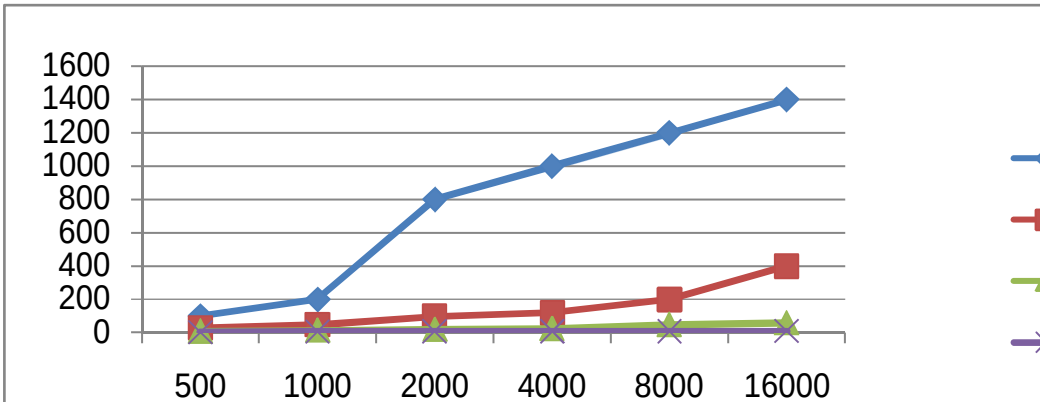
F. Relevance feedback:
Relevance feedback is method that used to improve the effectiveness of image retrieval. The main aim is to get the positive and negative results for a retrieved image from the user. For a given query, the system first retrieves a list of ranked images according to a predefined similarity metrics. When user get relevant image for a given query then user gives positive feedback otherwise negative feedback from user.based on user feedback system performance will br improved.

G. Retrieval results:
Image retrieval system mainly based on the user interaction. Images that are retrieved mainly based on user query and query results can be refined based on the relevance feedback of users. By getting feedback from user results can be modified finally get the exact result .

Algorithm:
simlearn algorithm
1. Initialization;
2. Iterate f
3. For images, $Sm+1(i; j) = \alpha Gm(i; j) + \beta Tm(i; j)$;
4. Feature learning to update $W = W*m+1$;
5. Update $S_{m+1}(i; j) = (1-\Phi cij)+\Phi ijS_{m+1}(i; j)$;

6. For groups, $S_{m+1}(g; g') = \alpha_G S^I_m(g; g') + \beta_G S^T m(g; g')$;
7. For tags, $S_{m+1}(t; t') = \alpha_G S^I_m(t; t') + \beta_G S^T_m(t; t0)$;
8. Until converge or stop criteria satisfied.

Time Efficiency:

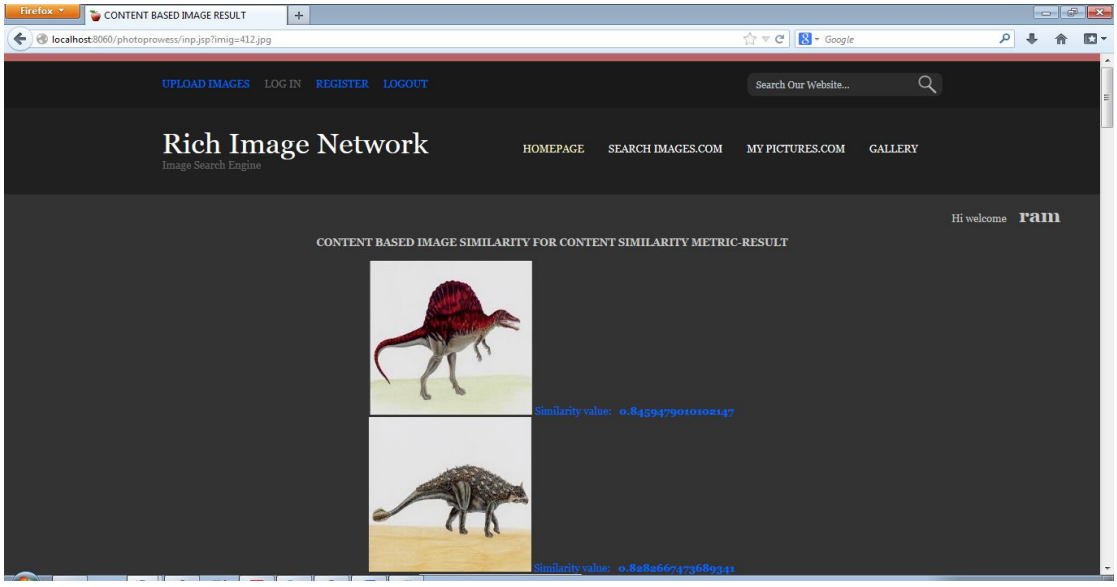


Time performance, X-axis denotes the number of images; Y-axis denotes the running time in seconds.

Results

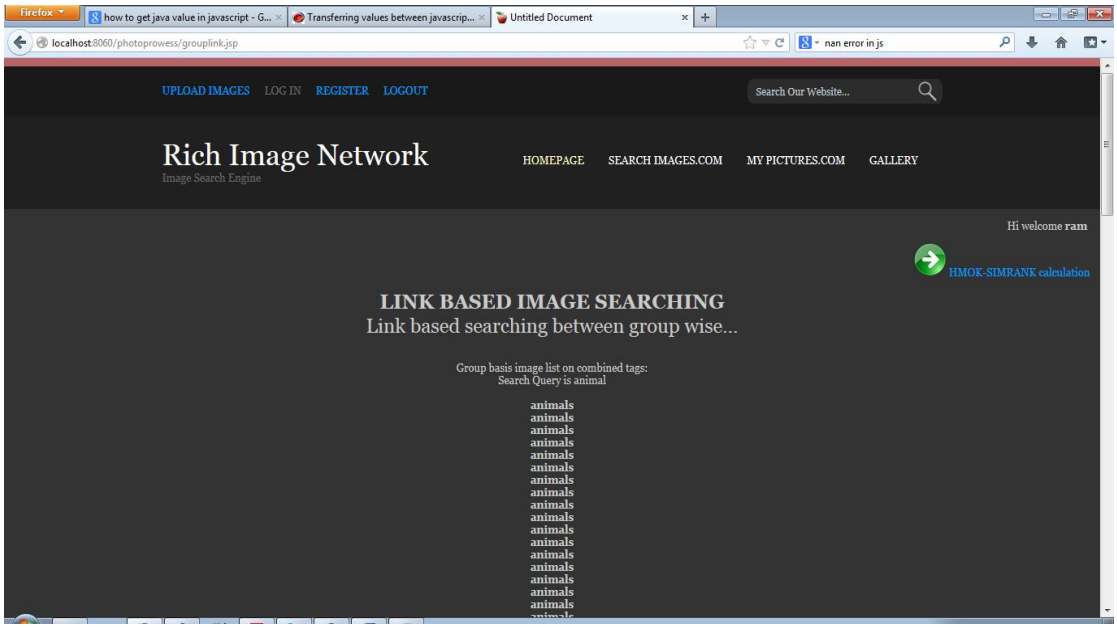
Content based image similarity:

Here it shows the result **Content based image similarity** based on features of image. When user enters the query it performs similarity based on the features of images.



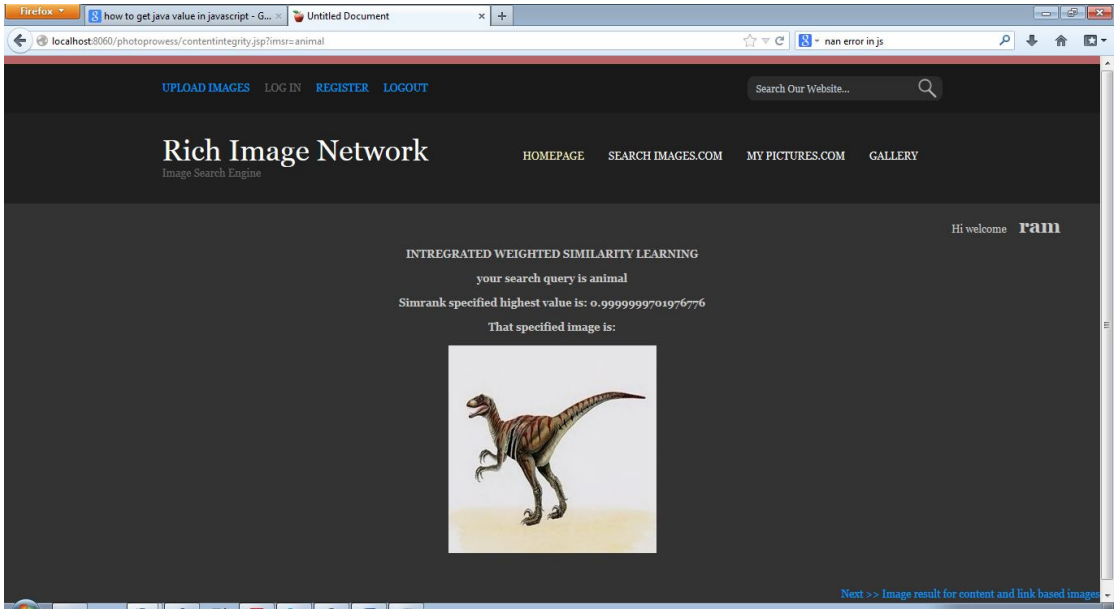
Link based similarity search:

Applications like social networks and biological databases having graphs like structure .we can understand the relationship between the nodes with the help of link prediction, recommendations and spam detection methods .with the help of link based similarity join we can find the link based similarities



Integration search:

In integration search which combines both link based and content base similarites using sim learn technique.



IV. Conclusion

Sim Learn is very helpful to find similar images, tags and groups. It provides better performance comparing to sim rank algorithms. The details are beyond the scope of this demo paper.

Sim Learn obtains the most relevant matches in both semantics and visual appearances. It correctly finds all the similar groups, for example, the groups similar to group "FLOWERS" are Flower Photography, FLOWER POWER" and Flower Pictures .

Scope of future enhancement:

To use integrated weight similarity learning in web search engine a distributive computing expansion will be investigated. There is an optimal solution with two steps, first perform the network partition that will divide large network into sub networks. Second, perform the simlearn techniques on each sub networks. Then combine all sub networks results to find the image similarity.

References

[1] R.L. Cilibrasi and P.M.B. Vitanyi, “The GooglSimilarityDistance,” IEEE Trans. Knowledge and Data Eng., vol. 19, no. 3,pp. 370-383, Mar. 2007

[2] L. Wu, X.-S. Hua, N. Yu, W.-Y. Ma, and S. Li, “Flickr Distance,”Proc. 16th ACM Int’L conf. Multimedia, pp. 31-40, 2008.

[3] Larry S. Davis, Rogerio S. Feris“ Image Ranking and Retrieval based on Multi-Attribute Queries” .

[4] Denis Turdakov and Dmitry Lizorkin “ Accuracy Estimate and Optimization Techniques for Sim Rank Computation”

[5] R.C. Veltkamp and M. Tanase, “Content-Based Image Retrieval Systems: A Survey,” technical report, Dept. of Computing Science, Utrecht Univ., 2002.

[6] H. Tamura and N. Yokoya, “Image Database Systems: A Survey,” Pattern Recognition, vol. 17, no. 1, pp. 29-43, 1984.

[7] Y. Rui, T.S. Huang, and S.-F. Chang, “Image Retrieval: Current Techniques, Promising Directions, and Open Issues,” J. Visual Comm. and Image Representation, vol. 10, no. 1, pp. 39-62, 1999.

[8] R. Datta, D. Joshi, J. Li, and J.Z. Wang, “Image Retrieval: Ideas, Influences, and Trends of the New Age,” ACM Computing Surveys, vol. 40, no. 2, pp. 1-60, Apr. 2008.